

Attachment 227

FEDERAL TRADE COMMISSION,

Plaintiff,

v.

AMAZON.COM, INC., a corporation;

NEIL LINDSAY, individually and as an
officer of AMAZON.COM, INC.;

RUSSELL GRANDINETTI, individually and
as an officer of AMAZON.COM, INC.; and

JAMIL GHANI, individually and as an officer
of AMAZON.COM, INC.,

Defendants.

Case No. 2:23-cv-0932-JHC

Honorable John H. Chun

REBUTTAL EXPERT REPORT OF DR. RAN KIVETZ (CORRECTED)

CONFIDENTIAL

whether presenting an Amazon customer with a Prime upsell had a meaningful likelihood of yielding a potentially unintentional enrollment. When considering the [REDACTED] rate at which upsells lead (or “convert”) to any enrollment in Prime, and the [REDACTED] rate at which Prime members (i) cancel, (ii) choose to take the survey, (iii) select “DNI,” and (iv) do not exhibit behaviors that are inconsistent with unintentional enrollment, I estimated the upper bound at which upsells result in potentially unintentional enrollments to be *approximately* [REDACTED].

132. My upsell analysis shows that, within the relevant context of the customer journey, the challenged upsells were extremely unlikely to result in unintentional enrollment (e.g., such that it would take on average [REDACTED] upsells to result in a single potentially unintentional enrollment). In particular, the results of my upsell analysis indicate that Amazon customers were consistently and readily able to decline offers to join Prime. Unlike scenarios where most consumers may not even be exposed to a particular communication (such as “junk” and “phishing” emails, mail, phone calls, text and voice messages, or other communications, which are often not even viewed/heard by the consumer),²³¹ Amazon customers *were* exposed to Prime upsells (including the at-issue upsells presented during the checkout process), and the very high rates at which customers declined these offers cannot be explained by simply not encountering or considering them.

133. Overall, the very low rates of what could even be argued to be consistent with unintentional enrollment, as shown through multiple pieces of evidence and analyses that Dr. Mahoney fails to consider, refute his assumption that “DNI” reliably measures unintentional enrollment and further invalidate his analysis approach.

B.1.4. *Amazon’s Business Practices Do Not Validate the Use of the Cancellation Survey to Measure Unintentional Enrollment*

134. Like Dr. Violette’s review of the Search Sentiment Survey, Dr. Mahoney evaluated the Cancellation Survey using incomplete criteria for assessing the reliability and/or validity of survey questions. In this and the next subsection, I explain why the factors identified

²³¹ Of course, unlike Prime, such “junk” communications (some of which consist of fraudulent and malicious cyberattacks, such as phishing messages) would also not enroll consumers into a benefits-like program after which, when canceled, many consumers would then report being satisfied with that program.

by Dr. Mahoney are insufficient to conclude that the Cancellation Survey provides reliable, accurate, and valid data on whether a customer unintentionally enrolled in Prime.

135. As one basis for relying on the results of the Cancellation Survey (and, in particular, on “DNI” selections), Dr. Mahoney points to the fact that Amazon employees used this survey as a source of “business intelligence.”²³² Whether Amazon derived insights from the survey and whether those insights informed its business practices is irrelevant, from an independent scientific perspective, to determining whether the survey can be used to reliably estimate or extrapolate the number of unintentional enrollees. As a customer satisfaction exit interview, the Cancellation Survey may very well have conferred—and, based on my understanding and review of relevant documents, did in fact confer²³³—relevant information in ways completely unrelated to measuring unintentional enrollment.

136. Regardless, any use by Amazon of the survey to derive business insights about its customers (including about why customers cancel) would *not* justify taking the survey’s “DNI” rates at face value and does *not* render the survey a scientifically reliable means of measuring unintentional enrollment. As I explained in Subsections A.2 and A.3 of my February 24, 2025 *Expert Report*, the Cancellation Survey may have been appropriate and useful from Amazon’s perspective in getting directional input about customer satisfaction and experiences with Prime, but it is *not* appropriate for measuring intentions and perceptions at the time of enrollment.

137. As Dr. Mahoney himself noted: “Like any empirical data, survey evidence should be assessed for its reliability for answering the research questions at hand.”²³⁴ The Cancellation Survey was neither designed to nor can it reliably answer the FTC’s research question in this litigation regarding the rate of unintentional enrollment in Prime. **If anything, the fact that the survey was conducted in the ordinary course of business—as primarily a customer satisfaction exit interview to gauge directional trends, *not* to quantify and extrapolate the rate of unintentional**

²³² *E.g.*, Mahoney Report, p. 29; *see also* ¶ 64.

²³³ *See* Subsection A.2.1 of my February 24, 2025 *Expert Report*.

²³⁴ Mahoney Report, ¶ 60.

enrollment²³⁵—is exactly a reason why the survey cannot simply be applied, as Dr. Mahoney does, to answer specific distinguishable questions in litigation, such as quantifying and extrapolating the rate of unintentional enrollment (and of “harm” due to alleged unintentional enrollment).

B.1.5. *The “Standard Survey Practices” that Dr. Mahoney Claims Were Used in the Cancellation Survey Do Not Validate its Use to Measure Unintentional Enrollment, and Dr. Mahoney Ignores Important Criteria that Invalidate the Survey for This Purpose*

138. Dr. Mahoney contends that the Cancellation Survey “followed standard survey practices,” specifically pointing to: “random sampling,” “randomized ordering of answers,” and gathering “a large number of responses from the population of interest.”²³⁶ While these are certainly rudimentary practices that surveys should follow, they do not constitute or capture the essence of what renders an instrument valid and reliable, and are therefore completely insufficient for determining whether a survey provides an appropriate basis to draw specific inferences about consumer behavior. Surveys that happen to meet Dr. Mahoney’s insufficient list of criteria can nevertheless be fatally flawed, leading, biased, and *irrelevant* to the research question at hand.

139. While Dr. Mahoney chose to focus on a subset of surface-level, “technical” survey practices, he ignored other material substantive flaws and issues with the Cancellation Survey that render its results scientifically unreliable. Such flaws relate to certain fundamental principles that surveys must follow not only to be sufficiently reliable as evidence for both academic and litigation purposes in general, but also to be sufficiently reliable with respect to assessing the FTC’s specific allegations about unintentional enrollment.

140. **The Mahoney Report ignores the fundamental ambiguity of the “DNI” answer choice.** Dr. Mahoney ignores the fact that the “DNI” answer choice he relies on—“I did not intend to sign up for Prime”—is overly broad and ambiguous. In asserting that the Cancellation Survey’s “questions were clear and non-leading, addressing concerns about demand effects,”²³⁷ Dr. Mahoney refers only to: the randomized order in which “pre-populated”²³⁸

²³⁵ See, e.g., AMZN_00037413; AMZN_00021537; AMZN_00001517; AMZN_00003654; AMZN_00040674; Hills Deposition, pp. 53, 55, 70, 181, 183, & 186.

²³⁶ Mahoney Report, p. 8.

²³⁷ *Id.*, ¶ 67.

²³⁸ *Ibid.*

B.3. The “DNI” Rates Across Nine Earlier Versions of the Cancellation Survey Do Not Validate, and Instead Undermine, Dr. Mahoney’s Assumption that the “DNI” Answer Choice Reliably Measures Unintentional Enrollment

166. Dr. Mahoney also attempts to justify his reliance on “DNI” selections in the 2020 Cancellation Survey—the longest-running version of the Cancellation Survey²⁷⁰—by pointing to purportedly analogous “DNI” rates from nine earlier survey versions. Figure B5 below reproduces the Mahoney Report’s Figure 22, which tabulates the “share,” or proportion, of responses that—according to Dr. Mahoney—indicate alleged unintentional enrollment in each of these nine other surveys.²⁷¹

Figure B5: Share of “DNI” Rates Across Nine Earlier Versions of the Cancellation Survey as Tabulated in Figure 22 of the Mahoney Report ²⁷²

Survey No.	Wording	Dates	Responses	Share
1	I did not mean to sign up for Amazon Prime	03/27/2020–05/04/2020		
2	I did not remember I signed up for Prime until I got charged	11/07/2018–03/02/2019		
3	Experienced a sign-up or cancellation problem with Amazon Prime	09/17/2018–03/31/2020		
4	I did not mean to sign up for Amazon Prime	11/07/2018–02/17/2019		
5	I did not mean to sign up for Amazon Prime	11/06/2018–03/05/2019		
6	I did not mean to sign up for Amazon Prime	11/06/2018–03/01/2020		
7	I did not mean to sign up for Amazon Prime	11/06/2018–04/26/2019		
8	I did not mean to sign up for Amazon Prime	04/03/2020–05/06/2020		
9	I did not mean to sign up for Amazon Prime	11/06/2018–04/02/2019		

167. *First*, the fact that the proportions of respondents who chose the “I did not mean to sign up for Amazon Prime” answer choice (what Dr. Mahoney describes as “the most similarly

²⁷⁰ I understand that in mid-2020, Amazon standardized the Cancellation Survey and, although there have been some changes, the version of the Cancellation Survey that launched in April 2020 in the United States has continued running and remained largely unchanged since that time. *See, e.g.*, AMZN_00039590. Between May 2020 and February 2024, approximately [REDACTED] customers responded to the 2020 Cancellation Survey. *See*

“SV_00InYvHGGzsKsBfUS+Prime+Cancellation+Survey+2020_February+5,+2024_16.13.csv.”

²⁷¹ Note that six of these “prior” survey versions (and particularly Survey #1) included data from respondents who started the survey after May 8, 2020, the date of the first recorded response in the 2020 Cancellation Survey. In his Figure 22, Dr. Mahoney truncated the data for the nine “prior” surveys, such that only data before May 8, 2020 were included. For ease of comparison, the analyses I report in this subsection are based on the same truncated data for the nine prior surveys (following Dr. Mahoney’s approach), although my conclusions remain the same when considering the entire, non-truncated data.

²⁷² Mahoney Report, ¶ 80.

worded option”²⁷³ to the “DNI” answer in the 2020 Cancellation Survey) were overall higher compared to the “DNI” rate in the 2020 Cancellation Survey does *not* imply that “it is reasonable if not conservative”²⁷⁴ to rely on the 2020 Cancellation Survey to estimate unintentional enrollment. For the reasons I discussed in Section A of my February 24, 2025 *Expert Report* (e.g., the time gap between enrollment and cancellation; the non-representativeness of the sample with respect to all Prime cancelers; the reliance on closed-ended, leading questions; the failure to minimize respondent guessing; and the lack of any survey control), my conclusions regarding the fundamental inability of the data from the 2020 Cancellation Survey to reliably measure alleged unintentional enrollment apply to these prior survey versions as well.

168. In particular, the answer choice “I did not mean to sign up for Amazon Prime” suffers from the same interpretation problems, including the use of overly broad and ambiguous wording, as the analogous “I did not intend to sign up for Prime” answer choice in the 2020 Cancellation Survey. A respondent taking one of these earlier surveys could plausibly have selected “I did not mean to sign up for Amazon Prime” for many reasons unrelated to unintentional enrollment, including because they did not mean to sign up for a *paid* subscription to Amazon Prime; because they did not mean to sign up for an ongoing, continuous subscription to Prime; or because *they* themselves did not mean to sign up for Prime (rather, someone else did).

169. **Second**, as Figure B5 makes clear, the “DNI” rates across these nine prior versions are unstable and vary *substantially* in magnitude—not only compared to each other but also compared to the [REDACTED] “DNI” rate in the 2020 Cancellation Survey (i.e., corresponding to the proportion of respondents in that survey who selected “DNI” in either Question 3 or 4 during the applicable period).²⁷⁵ Such discrepancies and variation, which the Mahoney Report fails to

²⁷³ *Ibid.*

²⁷⁴ *Ibid.*

²⁷⁵ Note that Dr. Mahoney’s analyses treated as unintentional enrollment only responses of “DNI” in Questions 3 or 4 (the “cancellation reasons” questions) of the 2020 Cancellation Survey. The overall “DNI” rate across Questions 3, 4, and 7 (which asked about problems experienced with Prime) was [REDACTED], calculated as [REDACTED] respondents out of a total of [REDACTED] respondents who took the survey between May 8, 2020 (the earliest recorded survey) and June 21, 2023 (the end date of the relevant period). Based on “SV_00InYvHGGzsKsBf US+Prime+Cancellation+Survey+2020_February+5,+2024_16.13.csv.”

acknowledge, show how sensitive the results are to different survey instruments and consumer samples—even when the identical question wording is used and when the surveys are administered during overlapping periods. For example, as Figure B5 above shows, Surveys #5 and 6 used an identical phrasing for the “DNI” option (“I did not mean to sign up for Amazon Prime”) and were administered during overlapping times (*i.e.*, between November 2018 and March 2019), yet generated dramatically divergent rates: [REDACTED] versus [REDACTED]—a [REDACTED] point difference that is itself greater than the [REDACTED] observed “DNI” rate in the 2020 Cancellation Survey.

170. Inexplicably, as discussed in the next subsection, Dr. Mahoney draws strong inferences from a small (only [REDACTED] point) difference in “DNI” rates when attempting to draw inferences about the purported effects of UPDP interface changes, yet chooses to ignore the much greater (up to nearly [REDACTED] point) differences in rates across different versions of the Cancellation Survey.

171. To further quantify the large differences, and instability, in “DNI” rates across the 10 Cancellation Survey versions (including the 2020 Cancellation Survey), I analyzed the degree of variance or heterogeneity, as a statistical matter, in these rates. Table B2 below shows the absolute value differences in “DNI” rates between all pairs of surveys, after excluding Surveys #4 and 8 due to their small sample sizes and including Survey #10 (the 2020 Cancellation Survey). These differences range in magnitude from a difference of [REDACTED] (*i.e.*, Surveys #5 vs. #9, each with a [REDACTED] “DNI” rate) to a difference of [REDACTED] (*i.e.*, Surveys #3 vs. #6, with “DNI” rates of [REDACTED] vs. [REDACTED] respectively). Out of 28 total pairwise comparisons, 19 comparisons (or 68%) show statistically significant differences, indicating *considerable* variation in “DNI” rates across the surveys being compared.²⁷⁶

²⁷⁶ Based on a Bonferroni-corrected χ^2 (“chi-square”) analysis. A χ^2 test is a statistical test that is commonly used to assess whether the observed frequencies or differences between sets of data with categorical (binary) variables significantly differ from expected frequencies (*i.e.*, over and above chance). Applying the Bonferroni correction is a conservative approach that adjusts the statistical significance threshold to control for Type I error (*i.e.*, “false positives”) when multiple comparisons are made. The standard in consumer behavior, marketing, communications, psychology, and other scientific research is that a statistic is considered statistically significant when its corresponding *p*-value is smaller than a significance level of .05 (*i.e.*, a less than 5% probability of a difference this large occurring by chance).

Table B2: Absolute Value Pairwise Differences in “DNI” Rates Across Cancellation Survey Versions²⁷⁷

	Survey 1 ████)	Survey 2 ████	Survey 3 ████	Survey 5 ████	Survey 6 ████	Survey 7 ████	Survey 9 ████	Survey 10 ████
Survey 1 ████)								
Survey 2 ████								
Survey 3 ████								
Survey 5 ████								
Survey 6 ████								
Survey 7 ████								
Survey 9 ████								
Survey 10 ████								

172. The results are similar when comparing only those surveys that used the *identical* wording of the “DNI” answer choice (“I did not mean to sign up for Amazon Prime”), as summarized in Table B3 below.²⁷⁸ Out of 10 pairwise comparisons, 8 differences—ranging from █████ to █████—are statistically significant.

(Continues on next page)

Under the Bonferroni correction, in this case, the *p*-value threshold for significance is set instead at $0.05 / 28 = .0018$.

²⁷⁷ I highlighted in orange the pairwise differences that are statistically significant (using the Bonferroni correction).

²⁷⁸ Again, after excluding Surveys #4 and 8 due to their small sample sizes, and after applying the Bonferroni correction.

Table B3: Absolute Value Pairwise Comparison Differences in “DNI” Rates Across Cancellation Surveys with the Same Answer Choice (“I Did Not Mean to Sign Up For Amazon Prime”) ²⁷⁹

	Survey 1	Survey 5	Survey 6	Survey 7	Survey 9
Survey 1 (████)	[REDACTED]				
Survey 5 (████)					
Survey 6 (████)					
Survey 7 (████)					
Survey 9 (████)					

173. To quantify the overall degree of heterogeneity or variation across surveys, I conducted an I^2 (“I-squared”) analysis by calculating the I^2 index—a useful heterogeneity measure ranging between 0% and 100% of the degree of (in)consistency between studies’ results,²⁸⁰ with I^2 values on the order of █████ and █████ being considered a low, moderate, and high degree of heterogeneity, respectively. The calculated I^2 index based on comparing the (analogous) “DNI” rates as tabulated by Dr. Mahoney for Surveys #1, #2, #3, #5, #6, #7, #9, and #10 was **99.8%**²⁸¹—corresponding to an *extremely high* degree of heterogeneity or variability across studies. Hence, the considerable spread in proportions is much larger than would be expected by chance,

²⁷⁹ I highlighted in orange the pairwise differences that are statistically significant (using the Bonferroni correction).

²⁸⁰ See, e.g., Higgins, Julian P.T., Simon G. Thompson, Johnathan J. Deeks, and Douglas G. Altman (2003), “Measuring Inconsistency in Meta-Analyses,” *BMJ: British Medical Journal*, 327(7414), 557 – 560; Borenstein, Michael, Larry V. Hedges, Julian P.T. Higgins, and Hannah R. Rothstein (2009), *Introduction to Meta-Analysis*, West Sussex, UK: Wiley. The I^2 index, first published in a widely-cited article, allows the researcher to calculate, even for a small number of studies, the percentage of the total variation across studies that is due to heterogeneity (*i.e.*, systematic inconsistency) among the studies rather than due to chance (*i.e.*, within-study error). See Higgins et al. (2003), p. 558; Borenstein et al. (2009), p. 117.

²⁸¹ Even after excluding the 2020 Cancellation Survey (*i.e.*, “Survey #10”), which had by far the largest sample size, the I^2 index is still █████. Moreover, after excluding Survey #6 (which had the highest “DNI” rate of █████), the I^2 index computed across the remaining seven surveys—████—is still extremely large.

suggesting that substantial differences exist between these surveys.²⁸² Such differences—whether a function of different samples, survey instruments, or other aspects—go unacknowledged and unaddressed in the Mahoney Report, and regardless are inconsistent with the conclusion that the 2020 Cancellation Survey is valid or reliable.

174. **Third**, while Dr. Mahoney did not report any evaluation of the nine prior Cancellation Survey versions beyond tallying their “DNI” rates, a closer inspection of the underlying survey instruments further evinces why the “DNI” answer choice in the 2020 Cancellation Survey cannot be relied upon to quantify unintentional enrollment. Of particular note is what the Mahoney Report refers to as “Survey #3” (*see* Figure B5 above), which differed in multiple meaningful ways from all other variants. The purported analogous “DNI” answer choice in this survey, per Dr. Mahoney’s tabulation, was “Experienced a sign-up or cancellation problem with Amazon Prime,” selected by [REDACTED] of the [REDACTED] respondents in response to the “cancellation reasons” question (Q.2): “Which of the following best describe [*sic*] why you didn’t renew your Amazon Prime paid membership?”²⁸³

175. It is unclear why Dr. Mahoney included Survey #3 in his table (*see* Figure B5) as being supposedly supportive of the notion that “many customers were unintentionally enrolled in Prime,”²⁸⁴ given that a response of “Experienced a sign-up or cancellation problem with Amazon Prime” is *not* diagnostic whatsoever of alleged unintentional enrollment, and does not even resemble an answer that approximates or is similar to “I did not intend to sign up for Prime.”

176. Unlike the other Cancellation Survey versions, this particular survey version also did not directly embed a “DNI” (or analogous “DNI”) answer choice to the “cancellation reasons” question. Instead, according to the underlying data, Question 2 essentially functioned as a “filter”

²⁸² I also compared the observed “DNI” rates for the four surveys with sufficiently large sample sizes and the same “I did not mean to sign up for Amazon Prime” answer choice (Surveys #5, #6, #7, and #9), holding constant the period of survey administration. Specifically, these four surveys shared an overlapping time period of November 7, 2018 to March 5, 2019, during which the “DNI” rates for Surveys #5, #6, #7, and #9 were [REDACTED] and [REDACTED], respectively—again revealing an extremely high degree of heterogeneity ($I^2 = 99.8\%$).

²⁸³ *See* “SV_2beUgzULi0i4ZA9 US+Prime+Cancellation+Survey+-+October+(ENDED)_February+5,+2024_16.32.csv.”

²⁸⁴ Mahoney Report, p. 42.

question²⁸⁵ for answers that could potentially reflect unintentional enrollment; more specifically, only those who selected the “Experienced a sign-up or cancellation problem with Amazon Prime” in Question 2 were asked in a subsequent question (Q.6): “Which of the following best explains what problems you experienced during sign up or cancellation with Amazon Prime?”²⁸⁶

Respondents could then choose one or more answers from among the following options:²⁸⁷

- I signed up for Amazon Prime by mistake
- I cancelled my Amazon Prime membership by mistake
- Amazon cancelled my account by mistake
- Amazon charged my card and signed me up without my consent
- I was not aware I had signed up for Amazon Prime until after it had happened
- Others (please specify): [text box]

177. Setting aside the methodological issues inherent in the above survey question (including, for example, a highly incomplete set of answers and the lack of any survey control), only [REDACTED] and [REDACTED] of respondents selected the two answer choices that come closest to articulating a potential instance of true unintentional enrollment: “Amazon charged my card and signed me up without my consent” and “I was not aware I had signed up for Amazon Prime until after it had happened” options, respectively. The proportion of respondents who selected *either* of these options was only [REDACTED]—a considerably lower rate than the [REDACTED] “DNI” rate observed in the 2020 Cancellation Survey, and without accounting for “noise” due to guessing and other survey flaws. Indeed, although the “I signed up for Amazon Prime by mistake” answer choice cannot be treated as diagnostic of unintentional enrollment,²⁸⁸ even if one were to count it as a *third* option

²⁸⁵ Leading survey authorities agree that filter questions reduce participant guessing and yea-saying, and assess whether participants have any opinions, perceptions, or beliefs *in the first place* about the specific topics raised in a question (e.g., their recollection of enrolling or their intention to enroll). See Subsection A.4.2 of my February 24, 2025 *Expert Report* for a discussion of the importance of filter questions and why the absence of any filter questions was problematic in the 2020 Cancellation Survey.

²⁸⁶ See “SV_2beUgzULi0i4ZA9 US+Prime+Cancellation+Survey+-+October+(ENDED)_February+5,+2024_16.32.csv.”

²⁸⁷ *Ibid.*

²⁸⁸ This answer choice is particularly broad and ambiguous, as signing up “by mistake” could very well refer to conscious actions taken by the customer when signing up for Prime—actions that the customer then perceived retroactively as a “mistake” and for which they blame themselves (*i.e.*, as something they should not have done, for any number of reasons unrelated to unintentional enrollment, such as over-estimating their future benefit usage).

(out of only five substantive options) potentially consistent with unintentional enrollment,²⁸⁹ this would still yield a rate of only [REDACTED] less than one third of the [REDACTED] that Dr. Mahoney opines is the predicted rate of unintentional enrollment.

178. Thus, consistent with my evaluation of the Cancellation Survey’s design in Section A of my previous report, the results of Survey #3 provide empirical evidence that correcting for only two flaws—namely, by employing a “filter” question and by using less ambiguous language (in both the question and the answer choices)—directly lead to significantly lower rates of responses that could indicate *potential* instances of unintentional enrollment.

179. **Overall**, Dr. Mahoney’s own tabulated summary of “DNI” rates across multiple versions of the Cancellation Survey invalidates, rather than supports, his reliance on the 2020 Cancellation Survey for purposes of measuring unintentional enrollment. The results of these nine prior surveys—which varied substantially even with the same question wording and during overlapping periods—provide further evidence that the “DNI” rates from the 2020 Cancellation Survey cannot be taken at face value, and, given the lack of any survey control, cannot be used to derive any reliable “net” rate of “DNI” selections, let alone alleged unintentional enrollment.

B.4. Dr. Mahoney’s Tabulation of Changes in “DNI” Rates During Certain Periods is Flawed, Anecdotal, and Does Not Show that the “DNI” Answer Choice Reliably Measures Unintentional Enrollment

180. In section III.A.4 of his report and as reproduced in Figure B7 below, Dr. Mahoney tabulates “DNI” rates in the 2020 Cancellation Survey over five time periods defined by what he claims to be changes in Amazon’s UPDP upsell user experience interface. He concludes from his tabulation the following:²⁹⁰

In general, during periods in which Amazon implemented clarity or user experience (“UX”) improvements in the UPDP upsell, the Cancellation Survey generates fewer “did not intend” responses. This has two implications. First, it provides support for the reliability of the Survey with respect to unintentional enrollment rates. Second, it provides evidence that less clarity on the Amazon UPDP upsell results in higher rates of unintentional enrollments.

²⁸⁹ [REDACTED] of respondents selected the “I signed up for Amazon Prime by mistake” answer choice.

²⁹⁰ Mahoney Report, ¶ 82.

Figure B7: “DNI” Rates Among UPDP Enrollees Based on their Enrollment Time Period as Tabulated in Figure 23 of the Mahoney Report

Time Period of UPDP enrollment		“Did not intend” responses
Time Period 1 (sparse Survey data)	March 19, 2020 and prior	
Time Period 2 (no shipping-based upsells)	April 23, 2020–September 16, 2020	
Time Period 3 (2020 clarity improvement)	September 17, 2020–December 2, 2020	
Time Period 4 (return to pre-COVID)	December 3, 2020–January 31, 2022	
Time Period 5 (CX Satisfaction Changes)	February 10, 2022–June 20, 2023	

181. According to Dr. Mahoney, the rates of “DNI” are “sharply lower”²⁹¹ between two “groups” of periods:²⁹²

[...] during (1) the time period in which the FTC indicates that Amazon was not offering free two-day shipping upsells (time period 2) and (2) the time periods in which the FTC indicates that Amazon implemented clarity improvements (time periods 3 and 5). For these periods, the rate of ‘did not intend’ responses is between [REDACTED] and [REDACTED]. In contrast, for UPDP signups during the 14-month duration of time period 4—when Amazon returned to the pre-COVID UPDP format that includes shipping-based upsells but not clarity or CX Satisfaction improvements—the rate of ‘did not intend’ responses is [REDACTED].

182. **First**, Dr. Mahoney’s statement that the “DNI” rates during Time Periods 2, 3, and 5 range “between [REDACTED] and [REDACTED]” is incorrect, even when compared to his own calculations and tabulation. In particular, it is unclear where Dr. Mahoney derived the erroneous [REDACTED] figure from, given that it does not appear in his tabulation of results (*see* Figure B7 above), which instead indicates that the “DNI” rates during the three highlighted periods ranged from [REDACTED] (Time Period 3) to [REDACTED] (Time Period 5). Indeed, the input data used to generate the Mahoney Report’s tabulation²⁹³ indicate that the average “DNI” rate in Time Period 4 (described as a “return to pre-COVID”) was [REDACTED] and that the average “DNI” rate in Time Period 5 (described as “CX Satisfaction Changes”) was [REDACTED]—a difference of only [REDACTED] points—which is not “sharply lower.”

²⁹¹ *Id.*, ¶ 88.

²⁹² *Ibid.*

²⁹³ Produced as “UPDP timeline.xlsx.”

183. In essence, Dr. Mahoney’s approach amounts to a comparison between the weighted average of “DNI” rates across Time Periods 2, 3, and 5 (*i.e.*, [REDACTED], reflecting the fact that, according to Dr. Mahoney’s calculations, the vast majority of the data are contained in Time Period 5) versus the corresponding “DNI” rate in Time Period 4 (*i.e.*, [REDACTED]). Again, this is a difference of only [REDACTED] percentage points between (i) the periods during which Amazon purportedly implemented “clarity improvements”²⁹⁴ and (ii) the period during which Amazon purportedly “returned to its pre-COVID UPDP format that includes shipping-based upsells but not clarity or CX Satisfaction improvements.”²⁹⁵

184. **Second**, although the above [REDACTED] percentage point difference is statistically significant, it is not *meaningful* for purposes of inferring and estimating alleged unintentional enrollment. As an initial matter, because statistical significance is heavily influenced by sample size, even miniscule differences can be statistically significant at large enough samples, despite being practically trivial.²⁹⁶ The risk of erroneously attributing meaning to small yet statistically significant effects or changes is further exacerbated when the researcher merely relies on statistical significance, without conducting any rigorous empirical analysis with proper controls.

185. In fact, Dr. Mahoney’s reported difference in average “DNI” rates during periods with supposedly “clear” versus “unclear” UPDP upsells is quite similar to the difference in average “DNI” rates during periods with different versions of supposedly “*clear*” upsells. Specifically, when comparing against the [REDACTED] “DNI” rate in Time Period 3 (during which Amazon apparently implemented a “2020 clarity improvement”²⁹⁷ in UPDP upsells), each of the two other periods that ostensibly were also associated with greater “clarity” nevertheless had statistically significantly *higher* “DNI” rates. The average “DNI” rate in Time Period 2—during which Amazon was apparently not even offering *any* free two-day shipping UPDP upsells (upsells that, according to Dr. Mahoney, the FTC alleges as being “unclear”²⁹⁸)—was [REDACTED] higher (a

²⁹⁴ Mahoney Report, ¶ 88 & FN 83.

²⁹⁵ *Ibid.*

²⁹⁶ *See, e.g.*, Kaye and Freedman (2000), pp. 252 – 253.

²⁹⁷ Mahoney Report, ¶ 89 (Figure 23).

²⁹⁸ *Id.*, ¶ 84.

statistically significant difference at the 1% level; $p < .01$) than in Time Period 3, when Amazon reinstated free two-day shipping UPDP upsells but with a purported “clarity improvement.” Likewise, compared to Time Period 3, the average “DNI” rate in Time Period 5—during which Amazon also made a purported “clarity improvement” by implementing “CX Satisfaction Changes”—was ■ percentage points higher, a statistically significant difference far below the .01 significance level; $p < .00001$).²⁹⁹

186. The Mahoney Report does not opine on which “clarity improvements” implemented by Amazon were ostensibly “clearer,” let alone to what degree they were “clearer.” Hence, there is no basis for concluding that UPDP “clarity improvements” must have caused fewer selections of the “DNI” answer choice.

187. **Third**, Dr. Mahoney’s assertion that the variation in “DNI” rates between different time periods supports his “premise”³⁰⁰ that “Amazon’s clarity improvements explain the reductions in the rates of ‘did not intend’ responses”³⁰¹ is highly speculative and does not substantiate any such “premise.” His analysis approach is akin to a qualitative version of an “event study” that attempts to assess the effect of a specific event (here, changes in UPDP upsells) on an outcome (selections of “DNI” in the Cancellation Survey) by comparing data before and after occurrence of that event. However, unlike a proper event study, Dr. Mahoney’s approach is uncontrolled—meaning that it does not include any controls, covariates, and/or appropriate difference-in-differences analysis³⁰² to account for other factors (unrelated to “clarity” experiments or unintentional enrollment) that could explain the observed differences. Simply “eyeballing” trends is insufficient for deriving credible causal inferences about alleged unintentional enrollment.

188. **Fourth**, even taking at face value the Mahoney Report’s calculations of changes in “DNI” rates vis-à-vis changes in UPDP upsells, and even assuming for the sake of argument that “DNI” is associated with unintentional enrollment, the fact that Dr. Mahoney may have

²⁹⁹ The p -value associated with this ■ percentage point difference is approximately 6.7×10^{-22} , or effectively $p = 0$.

³⁰⁰ *Id.*, ¶ 89.

³⁰¹ *Ibid.*

³⁰² See, e.g., Angrist, Joshua D. and Jörn-Steffen Pischke (2009), *Mostly Harmless Econometrics: An Empiricist’s Companion*, Princeton, NJ: Princeton University Press.

found *some* sensitivity in “DNI” survey responses to changes in UPDP upsells would still not validate using an observed “DNI” rate to measure and extrapolate unintentional enrollments for purposes of damages estimation. Among other reasons, this is because the small differences in “DNI” rates (ostensibly due to variations in UPDP upsells), which Dr. Mahoney focuses on, stand in stark contrast to the *much larger* differences or variability in “DNI” rates due to other aspects that he ignores.

189. Consider again the nine prior versions of the Cancellation Survey (reproduced in Figure B8 below). Dr. Mahoney draws inferences from a small (*e.g.*, [REDACTED] point) difference over time in the “DNI” rate when drawing inferences about the effect of purported UPDP “clarity improvements,” yet chooses to overlook the much greater differences (of up to nearly [REDACTED] points) across Cancellation Survey variants—including those that use identical question wording. The fact that “DNI” rates vary so dramatically across survey versions undermines any reliance on a single survey instrument, particularly when the temporal trends within that single instrument are small in comparison to the between-survey variability.

Figure B8: Share of “DNI” Rates Across Nine Earlier Versions of the Cancellation Survey as Tabulated in Figure 22 of the Mahoney Report ³⁰³

Survey No.	Wording	Dates	Responses	Share
1	I did not mean to sign up for Amazon Prime	03/27/2020–05/04/2020	[REDACTED]	[REDACTED]
2	I did not remember I signed up for Prime until I got charged	11/07/2018–03/02/2019		
3	Experienced a sign-up or cancellation problem with Amazon Prime	09/17/2018–03/31/2020		
4	I did not mean to sign up for Amazon Prime	11/07/2018–02/17/2019		
5	I did not mean to sign up for Amazon Prime	11/06/2018–03/05/2019		
6	I did not mean to sign up for Amazon Prime	11/06/2018–03/01/2020		
7	I did not mean to sign up for Amazon Prime	11/06/2018–04/26/2019		
8	I did not mean to sign up for Amazon Prime	04/03/2020–05/06/2020		
9	I did not mean to sign up for Amazon Prime	11/06/2018–04/02/2019		

190. Contrary to the Mahoney Report’s reliance on “DNI” responses as representing unintentional enrollment, the results of my customer behavior analysis and my analysis of “internal inconsistencies” in the Cancellation Survey data indicate that the vast majority (*e.g.*, approximately

³⁰³ Mahoney Report, ¶ 80.

████) of respondents who selected “DNI” exhibited behaviors or provided responses *inconsistent* with unintentional enrollment. Because the “DNI” answer choice does not reliably measure the construct at issue, a high proportion of “DNI” respondents would be misclassified (*i.e.*, overclassified) by Dr. Mahoney as unintentional enrollees in each of his five periods. Further, the differences in the “DNI” rates across these periods are considerably smaller than—and overshadowed by—the rate of misclassification. The gross inflation in the “DNI” rate observed in the 2020 Cancellation Survey means that the existence of small fluctuations in the data (*e.g.*, a █████ percentage point difference) are irrelevant and do not justify computing an absolute or “net” rate of unintentional enrollment, as Dr. Mahoney did.

B.5. Dr. Mahoney’s Comparison of Benefit Usage Between “DNI” and Non-“DNI” Respondents is Flawed, Does *Not* “Corroborate” the Cancellation Survey Results or Show that “DNI” Reliably Measures Unintentional Enrollment, and, if Anything, Supports My Empirical Analyses

191. To support his position that “DNI” responses in the Cancellation Survey reliably indicate unintentional enrollment, Dr. Mahoney presents a figure showing that respondents who selected “DNI” used fewer Prime shipping benefits overall than did non-“DNI” respondents (reproduced as Figure B9 below). According to Dr. Mahoney, this pattern “is consistent with the idea that someone who is not aware of their Prime membership would use fewer benefits, thus corroborating the Survey results”;³⁰⁴ he further argues that “[t]o the extent the Survey is correctly reporting unintended enrollments, usage of Prime benefits should be lower for those who report unintentionally enrolling than for those who do not.”³⁰⁵ For the reasons below, this pattern does *not* in fact “corroborat[e]” the Cancellation Survey’s results or demonstrate that unintentional enrollment can be reliably inferred from “DNI” rates.

(Continues on next page)

³⁰⁴ *Id.*, ¶ 92.

³⁰⁵ *Id.*, ¶ 90.